1. Designing maps from texts

Semiology teaches us that maps as texts express ideas. In this thesis work which is still in progress, we focus on maps which come with texts. In this particular case, texts and maps are supposed to express the same ideas. The point is to help users to make map which conveys the same message than the text it accompanies.

A map may be seen as the result of the application of a key map to a geographic dataset. Although information about relevant dataset can be found in the text, this work only talks about the key map.

2. Extracting and structuring the text message

A message results from an intellectual construction which can be seen as structuring main topics and concepts pertaining to these topics. In order to extract the text message, tools from natural language processing and knowledge management are usually utilized. These techniques are based on corpora. Thus, a corpus of French texts from various sources (newspapers, magazines, etc.) and accompanied by maps has been collected. The corpus is specialized; its theme is the geopolitics of natural resources, assuming that the approach can be applied to others themes.

The first step of the text analysis consists in identifying the main topics. To complete this task, lexical statistics are used. For example, Lebart & Salem (1994) make it possible to identify the most frequent and the significantly frequent lemmas of a text, i.e. the relevant concepts of the specialized corpus, by Chi-square measures with a reference corpus. In this case, the reference corpus is non-specialized and journalistic. In the specialized corpus, the lemma *pétrole* (*oil* in English), *gaz* (*gas)* or *mine* (*mine*) has been identified.

The second step consists in extracting and structuring the concepts pertaining to the identified topics. This set of concepts comes from two types of resources: a field ontology containing concepts like *champ de pétrole* (*oil field)*, or *forage pétrolier* (*oil well*) and based on already existing ontologies (Horrocks 2008), and terminological resources which make it possible to identify the “semantic relations” (Bertin 1967) between thematic concepts. These text segments are retrieved thanks to a stream of morpho-syntactic patterns applied to the text. The result of the second step is a key map structure based on Jacques Bertin’s semantic relations.

3. Assigning a graphic representation to the identified topics

The assignment of a graphic sign to each line of the key map is done by applying a set of rules that has been defined beforehand to the key map structure. This set can be divided into two groups.

The first is related to Jacques Bertin's graphic semiology that deals predominantly with both semantic relations and visual variables. The first ones are called *selectivity*, *associativity*, *order* and *quantity* and have allowed to organize the key map structure through the structuring process mentioned above. Thus symbolization suggestions can be made on the basis of the visual variables, for example, using value variation to express order relation.

The second group of rules could be called the user’s preferences and is based on the cartographer’s daily practices guided by their experience. These rules are laid down after interview with some cartographers who explained how they design maps on the basis of a text. The specialized corpus has been analyzed in order to identify recurrent graphic signs. The latter are realistic symbols and symbolic colors, as well as representations that are usually assigned to design changes of geographical objects over time.

References

Bertin J (1967) Semiology of Graphics: Diagrams, Networks, Maps. University of Wisconsin Press, 1983 (first published in French in 1967, translated to English by Berg W.J. in 1983)

Horrocks I (2008) Ontologies and the semantic web. Communications of the ACM, 51(12):58–67

Lebart L & Salem A (1994) Statistique textuelle. Dunod, Paris